Chapter Review Exercises (page 664)

R10.1 (a) Paired t test for the mean difference.
(b) Two-sample z interval for the difference in proportions.
(c) One-sample t interval for the mean.
(d) Two-sample t interval for the difference between two means.

R10.2 (a) For these data, $\hat{p}_1$ = sample proportion of Hispanic female drivers in New York who wore seat belts = $\frac{183}{220} = 0.832$ and $\hat{p}_2$ = sample proportion of Hispanic female drivers in Boston who wore seat belts = $\frac{68}{117} = 0.581$. Thus,

$$\text{SE}_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{0.832(1-0.832)}{220}+\frac{0.581(1-0.581)}{117}} = 0.0521.$$

If we were to take many random samples of 220 Hispanic female drivers in New York and 117 Hispanic female drivers in Boston, the difference in the sample proportions from New York and Boston who wear seatbelts will typically be 0.0521 from the true difference in proportions of all Hispanic female drivers in New York and Boston that wear seat belts.
(b) *State:* Our parameters of interest are p_1 = proportion of all Hispanic female drivers in New York who wear seat belts and p_2 = proportion of all Hispanic female drivers in Boston who wear seat belts. We want to estimate the difference $p_1 - p_2$ at a 95% confidence level. *Plan:* We should use a two-sample z interval for $p_1 - p_2$ if the conditions are met. Random: The data come from independent random samples. 10%: $n_1 = 220$ is less than 10% of all Hispanic female drivers in New York and $n_2 = 117$ is less than 10% of all Hispanic female drivers in Boston. Large Counts: Both samples have at least 10 successes and failures (New York: 183 successes and 37 failures. Boston: 68 successes and 49 failures). *Do:* From the data we find that $n_1 = 220$, $\hat{p}_1 = \frac{183}{220} = 0.832$, $n_2 = 117$, and $\hat{p}_2 = \frac{68}{117} = 0.581$. The 95% confidence interval is

$$(0.832-0.581) \pm 1.96\sqrt{\frac{0.832(0.168)}{220}+\frac{0.581(0.419)}{117}} = 0.251 \pm 0.102 = (0.149, 0.353).$$

Conclude: We are 95% confident that the interval from 0.149 to 0.353 captures the true difference in the proportions of Hispanic women drivers in New York and Boston who wear their seat belts.

R10.3 (a) The Random condition is still met because the women in the study were randomly assigned to one of the two treatments.
(b) Because both groups are large ($n_C = 45 \geq 30$ and $n_A = 45 \geq 30$), the sampling distribution of $\bar{x}_C - \bar{x}_A$ should be approximately Normal.
(c) Assuming there is no difference in the true mean ratings of the product for women like these who read or don't read the news story, there is less than a 0.01 probability of observing a difference as large as or larger than the one observed in this experiment (5.05 – 4.56 = 0.49) by chance alone.

R10.4 (a) *State:* Our parameters of interest are μ_1 = the true mean NAEP quantitative skills test score for young men and μ_2 = the true mean NAEP quantitative skills test score for young women. We want to estimate $\mu_1 - \mu_2$ at the 90% confidence level. *Plan:* We should use a two-sample *t* interval for $\mu_1 - \mu_2$ if the conditions are met. Random: Even though the data came from a single random sample, it is reasonable to consider the two samples independent because knowing the response of a male shouldn't help us predict the response of a female. 10%: n_1 = 840 is less than 10% of all young men and n_2 = 1077 is less than 10% of all young women. Normal/Large Sample: n_1 =840 ≥ 30 and n_2 = 1077 ≥ 30. *Do:* The conservative degrees of freedom is 840 – 1 = 839. Using Table B and df = 100, the confidence interval is $(272.40-274.73)\pm 1.660\sqrt{\frac{(59.2)^2}{840}+\frac{(57.5)^2}{1077}}$ = –2.33 ± 4.47 = (–6.80, 2.14). Using technology: (–6.76, 2.10) with df = 1777.52. *Conclude:* We are 90% confident that the interval from –6.76 to 2.10 captures the true difference in the mean NAEP quantitative skills test score for young men and the mean NAEP quantitative skills test score for young women.

(b) Because 0 is in the interval, it is plausible that the true difference is 0. That is, we do not have convincing evidence of a difference in mean score for male and female young adults.

R10.5 (a) *State:* We want to perform a test at the $\alpha = 0.05$ significance level of $H_0: p_1 - p_2 = 0$ versus $H_a: p_1 - p_2 < 0$ where p_1 is the true proportion of patients like these who take AZT and develop AIDS and p_2 is the true proportion of patients like these who take placebo and develop AIDS. *Plan:* We should use a two-sample *z* test for $p_1 - p_2$ if the conditions are met. Random: These data come from two groups in a randomized experiment. Large Counts: The number of successes and failures in both groups are at least 10 (AZT: 17 successes, 418 failures. Placebo: 38 successes, 397 failures). *Do:* The proportions of AIDS cases in each group are $\hat{p}_1 = \frac{17}{435} = 0.039$ and $\hat{p}_2 = \frac{38}{435} = 0.087$. The pooled proportion is $\hat{p}_C = \frac{17+38}{435+435} = \frac{55}{870} = 0.063$. The test statistic is $z = \frac{(0.039-0.087)-0}{\sqrt{\frac{(0.063)(0.937)}{435}+\frac{(0.063)(0.937)}{435}}} = -2.91$. The *P*-value is 0.0018. *Conclude:* Because the *P*-value of 0.0018 is less than α = 0.05, we reject H_0. We have convincing evidence that taking AZT lowers the proportion of patients like these who develop AIDS compared to a placebo.
(b) A Type I error is finding convincing evidence that AZT lowers the risk of developing AIDS, when in reality it does not. A consequence is that patients will pay for a drug that doesn't help. A Type II error is not finding convincing evidence that AZT lowers the risk of developing AIDS, when in reality it does. A consequence is that patients won't take the drug when it could actually delay the onset of AIDS. Because we found convincing evidence that AZT lowers the risk of developing AIDS, it is possible that we made a Type I error.

R10.6 (a) The Large Counts condition is not met because there are only 7 failures in the control area.

(b) The Normal/Large Sample condition is not met because both sample sizes are small and there are outliers in the male distribution.

R10.7 (a) Even though each subject has two scores (before and after), the two groups of students are independent. That is, knowing the improvement of a subject in one group will not help predict the improvement of a student in the other group.
(b) The distribution of differences for the control group is slightly skewed to the right while the distribution of differences for the treatment group is roughly symmetric. The center for the treatment group is greater than the center for the control group. The differences in the control group are more variable than the differences in the treatment group. Overall, it appears that students in the treatment group had bigger improvements, on average. (c) *State:* We want to perform a test at the $\alpha = 0.05$ significance level of $H_0 : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 > 0$ where μ_1 = the true mean difference in test scores for students like these who get the treatment message and μ_2 = the true mean difference in test scores for students like these who get the neutral message. *Plan:* We should use a two-sample *t* test for $\mu_1 - \mu_2$ if the conditions are met. Random: The data come from two groups in a randomized experiment. Normal/Large Sample: Both samples had less than 30 observations, but neither boxplot showed strong skewness or any outliers, so it is reasonable to perform a two-sample *t* procedure. *Do:* Using the differences, we know that $n_1 = 10,\ \bar{x}_1 = 11.4,\ s_1 = 3.169,\ n_2 = 8,\ \bar{x}_2 = 8.25,$ and $s_2 = 3.69.$ The test statistic is $t = \dfrac{(11.4 - 8.25) - 0}{\sqrt{\dfrac{(3.169)^2}{10} + \dfrac{(3.69)^2}{8}}} = 1.91.$ Using the conservative degrees of freedom (8 – 1 = 7), the *P*-value is between 0.025 and 0.05. *Using technology*: *t* = 1.91, df = 13.92, *P*-value = 0.0382. *Conclude:* Because the *P*-value of 0.0382 is less than α = 0.05, we reject H_0. There is convincing evidence that the true mean difference in test scores for students like these who get the treatment message is greater than the true mean difference in test scores for students like these who get the neutral message.
(d) We cannot generalize to all students who failed the test because our sample was not a random sample of all students who failed the test. It was a group of students who agreed to participate in the experiment.